LU preconditioning for singular sparse least squares

Nick Gould, Michael Saunders, and Jennifer Scott

Rutherford Appleton Laboratory, UK
ICME and MS&E, Stanford University

SIAM Conference on Computational Science and Engineering (CSE17)
Atlanta, GA Feb 27 – Mar 3, 2017

Abstract

For overdetermined systems $Ax \approx b$ with full column rank, LU factors of A with L well-conditioned give U as a reasonable right-preconditioner, as implied by Peters and Wilkinson (1970). When A has low column rank, U is singular. We show that it still provides a helpful preconditioner for computing the minimum-length solution x for the original system. We experiment with LUSOL on a range of sparse singular problems.

Partially supported by the
National Institute of General Medical Sciences of the National Institutes of Health (NIH)
Award U01GM102098

Background

Direct and iterative methods for least squares

Dense $\min \|Ax - b\|$

- QR — full-rank (Golub 1965)
- QR — tall and skinny (MapReduce: Benson, Gleich & Demmel 2013)
- SVD — singular

Sparse $\min \|Ax - b\|^2 + \delta^2 \|x\|^2$

- SuiteSparseQR (Davis 2011) — full-rank or singular
- LSQR (Paige & S 1982) — same
- LSMR (Fong & S 2011) — same
- LSRN (Meng, S, & Mahoney 2014) — dense or sparse, tall and skinny

LU preconditioning

Dense $\min \|Ax - b\|$, full-rank

$PA = LU$ Partial Pivoting $|L_{ij}| \le 1$

Peters and Wilkinson 1970 $L^T L y = L^T P b, \quad Ux = y$

Sparse $\min \|Ax - b\|$, full-rank

$P_1 A P_2 = LU$ Threshold Partial Pivoting (TPP) $|L_{ij}| \le 1.1$

S, 1979	LSQR	preconditioner U, B	row-oriented TPP	
Arioli & Duff 2015	LSQR	preconditioner B	MA58	TPP, TRP
Today	LSMR	preconditioner U	LUSOL	TPP, TRP
			UMFPACK	TPP

LU preconditioning for $\min \|Ax - b\|$

Peters and Wilkinson 1970 Dense A

LU with Partial Pivoting

- $PA = LU \Rightarrow \|Ax - b\| = \|LUx - Pb\|$ — P is likely to keep L well-conditioned
- $\min \|Ax - b\| \equiv \min \|Ly - Pb\|$
- $L^T Ly = L^T Pb$ is ok
- $Ux = y$

Saunders 1979 sparse A

LSQR with preconditioner U

- $PA = LU = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} U = \begin{bmatrix} B \\ N \end{bmatrix}$ $|L_{ij}| \le 1.1 \Rightarrow L$ probably well-conditioned
- Primitive row-oriented LU
- LSQR improved on illc1033

Suggested LSQR with preconditioner B

- $\|Ax - b\| = \left\| \begin{bmatrix} B \\ N \end{bmatrix} x - Pb \right\| \quad \Rightarrow \quad \left\| \begin{bmatrix} I \\ NB^{-1} \end{bmatrix} y - Pb \right\|, \quad Bx = y$
- $NB^{-1} = L_2 L_1^{-1}$
- $\|NB^{-1}\|$ should not be large $\Rightarrow$ $\begin{bmatrix} I \\ NB^{-1} \end{bmatrix}$ should be well-conditioned
- Can do sparser $B = LU$ once B is found

Arioli & Duff 2015 sparse A

Modern sparse $PA = LU$ (column perm also)

- $PA = LU = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} U = \begin{bmatrix} B \\ N \end{bmatrix}$ $\quad |L_{ij}| \le 1.1 \Rightarrow L$ probably well-conditioned
- $\|Ax - b\| = \left\| \begin{bmatrix} B \\ N \end{bmatrix} x - Pb \right\| \quad \Rightarrow \quad \left\| \begin{bmatrix} I \\ NB^{-1} \end{bmatrix} y - Pb \right\|, \quad Bx = y$
- Focus on $\det(U) = \text{volume}(B)$ large
- $\|NB^{-1}\|$ should not be large $\quad \Rightarrow \quad \begin{bmatrix} I \\ NB^{-1} \end{bmatrix}$ should be well-conditioned
- B should be good preconditioner
- $Bx_0 = b_B$ gives starting point x_0

LSRN
Meng, S, & Mahoney, SISC (2014)
for $m \gg n$ or $m \ll n$

LSRN for $m \gg n$ or $m \ll n$

- Compress A with **random normal transformation** G
- Get preconditioner from QR or SVD of GA or AG
- Ok for singular systems

Random normal projection, $m \gg n$

- Choose $s = 2n$, say
- Define $G = \texttt{randn}(s, m)$
- Compute $GA = U\Sigma V^T$ (compact SVD, rank $r \le n$)
- Let $N = V\Sigma^{-1}$
- Compute min-length solution of $\min_y \|ANy - b\|$ (LSQR, CS)
- Return $x = Ny$

Preconditioner N from singular $A = LU$

Suppose $y^\dagger$ is the min-length solution of $\min_y \|ANy - b\|$.
We want $Ny^\dagger$ to be the min-length solution of $\min_x \|Ax - b\|$.

LSRN paper: LEMMA 3.1 $A^\dagger = N(AN)^\dagger$ iff $\text{range}(NN^TA^T) = \text{range}(A^T)$.
COROLLARY $x^\dagger = Ny^\dagger$ if $\text{range}(NN^TA^T) = \text{range}(A^T)$.

Suppose $A = LU$ with row/col permutations P, Q.

$$PAQ = PLP^T\ PUQ = \begin{pmatrix} L_1 & \\ L_2 & I \end{pmatrix} \begin{pmatrix} U_1 & U_2 \\ & 0 \end{pmatrix} = \begin{pmatrix} L_1 & \\ L_2 & 0 \end{pmatrix} \begin{pmatrix} U_1 & U_2 \\ & I \end{pmatrix}$$

Choose $N \equiv Q \begin{pmatrix} U_1 & U_2 \\ & I \end{pmatrix}^{-1} \Rightarrow AN = P^T \begin{pmatrix} L_1 & \\ L_2 & 0 \end{pmatrix}$ is singular but well-conditioned.

We can prove $\text{range}(NN^TA^T) = \text{range}(A^T) = \text{range} \begin{pmatrix} L_1^T & L_2^T \\ 0 & 0 \end{pmatrix}$. Hence $x^\dagger = Ny^\dagger$.

LSMR stopping tolerance

LSMR stops when $\dfrac{\|A^T r_k\|}{\|A\| \|r_k\|} \le \texttt{atol}$

LSMR stopping tolerance `atol`

Problem `tomographic1`: 73159×59498, 647495 nonzeros
Columns of A normalized $\Rightarrow$ Diagonal preconditioning $\qquad b = (1\ 1\ 1\ \dots\ 1)^T$

Matlab LSMR, Matlab data

`atol`	LSMR itns	$\|r_k\|$
1e−0	0	2.705e+02
1e−6	3000	1.728e+02
1e−7	9000	1.722e+02
1e−8	30000	1.718e+02
1e−9	101000	1.717e+02
1e−10	413000(!)	1.717e+02

Time 2080 secs

LSMR stopping tolerance `atol`

Problem `tomographic1`: 73159×59498, 647495 nonzeros
Columns of A normalized $\Rightarrow$ Diagonal preconditioning $\qquad b = (1\ 1\ 1\ \dots\ 1)^T$

f90 LSMR, RB data

atol	LSMR itns	$\lVert r_k \rVert$
1e-0	0	2.705e+02
1e-6	3000	1.785e+02
1e-7	8400	1.779e+02
1e-8	32000	1.774e+02
1e-9	109000	1.772e+02
1e-10	424000(!)	1.771e+02

Time 1630 secs

LUSOL
TPP, TRP, TCP
Threshold partial/rook/complete pivoting

Needed for MINOS, SNOPT, PATH, . . . basis handling

Also for $\begin{bmatrix} B & S \end{bmatrix} P = \begin{bmatrix} \bar{B} & \bar{S} \end{bmatrix}$

via TPP on $\begin{pmatrix} B^T \\ S^T \end{pmatrix}$ **keeping** L **well-conditioned**

LU with threshold pivoting keeping L and/or U well-conditioned

Best to think of

$$PAQ = LDU \qquad \text{like } A = U\Sigma V^T$$

$$L = \begin{bmatrix} L_1 & \\ L_2 & I \end{bmatrix} \qquad D = \begin{bmatrix} D_1 & \\ & 0 \end{bmatrix} \qquad U = \begin{bmatrix} U_1 & U_2 \\ & I \end{bmatrix}$$

L, U have unit diags

τ = threshold pivot tolerance (to preserve stability) $\tau = 1/u$

$\tau = 1.5$ here

TRP and TCP are rank-revealing

TPP	$\lvert L_{ij}\rvert \le \tau$	L well-conditioned
TRP	$\lvert L_{ij}\rvert \le \tau, \quad \lvert U_{ij}\rvert \le \tau$	L, U well-conditioned
TCP	$\lvert L_{ij}\rvert \le \tau, \quad \lvert U_{ij}\rvert \le \tau, \quad \lvert D_{11}\rvert \gtrsim \lvert D_{22}\rvert \gtrsim \dots$	L, U well-conditioned

U repair

- For all pivot options, PUQ is trapezoidal
- Solves with singular U and U^T pretend that $PUQ = \begin{pmatrix} U_1 & U_2 \\ & I \end{pmatrix}$
- `rank` $=$ no. of rows in $\begin{pmatrix} U_1 & U_2 \end{pmatrix}$
- Small diagonals of U_1 are changed to 1 (if $|U_{jj}| \le$ `Utol2 = 1e-6`)
- `nrepair` $=$ number changed

Preconditioning with singular U (TPP vs TRP)

Problem `Maragal_6`: 21255×10152, 537694 nonzeros

Columns of A normalized, $b = (1\ 1\ 1\ \ldots\ 1)^T$

$\tau = 1.5$ `Utol2` $=$ `1e-6`

`atol = 1e-08`

Precon	rank	nrepair	LSMR itns	Final $\|r_k\|$	Time
I			2241	94.0014	4.8
U, TPP	8331	3	failed	(huge nos)	
U, TRP	8335	4	408	93.8527	2.7

`atol = 1e-10`

Precon	LSMR itns	Final $\|r_k\|$	Time
I	18251	93.8586	38.
U, TPP	failed	(huge nos)	
U, TRP	569	93.8527	3.7

Our only success story!

Preconditioning with singular U from UMFPACK

Problem `Maragal_6`: 21255×10152, 537694 nonzeros
Columns of A normalized, $b = (1\ 1\ 1\ \ldots\ 1)^T$

```
thresh = 1/1.5;      [L,U,P,Q] = lu(A,thresh);
```

Luckily

- L is well-conditioned (if we add ϵ to last col of singleton rows)
- TPP (not TRP), but if diag(U) is very small, whole row of U is zero
- Multifrontal data structure $\Rightarrow$ faster than Markowitz LU

Unluckily

- Storage went up to 10G.
- Froze my 16G iMac.

Nick Gould and Jennifer Scott (RAL)

The state-of-the-art of preconditioners for sparse linear least-squares problems

ACM TOMS, Jan 2017

Gould and Scott 2017

- 921 rectangular problems from CUTEst and the UFL collection
- narrowed down to 83 problems for which LSMR needs more than 100,000 iterations or 10 seconds

Comprehensive experiments with

- direct methods on $C = A^TA + \delta I$, $K = \begin{pmatrix} I & A \\ A^T & -\delta I \end{pmatrix}$, $\delta = 10^{-12}, 10^{-10}$ resp.
- preconditioners for LSMR based on incomplete factorizations (IC, MIQR, RIF)
- stationary iterations (BA-GMRES)

Performance profiles

Recommend future comparisons on
PDE1, IMDB, GL7d17–21, NotreDame_actors, TF17–19, wheel_601

Conclusions, recommendations, prejudices, conjectures

- If A is an explicit sparse matrix (not a linear operator), normalize the columns before defining the problem: $A \leftarrow AD, \quad \min \|Ax - b\|, \quad x \leftarrow Dx$.
- Define $K = \begin{pmatrix} \alpha I & A \\ A^T & -\delta I \end{pmatrix}$ for some $\alpha \approx \sigma_r(A) < 1$.
- Then C, K don't need further scaling, α, δ, and stopping rules are more meaningful.
- `atol = 1e-6, 1e-7, ...` is critical for LSQR, LSMR. But try no-preconditioner first.
- G & S reverse-communication versions of LSQR, LSMR allow stopping tests on original problem (independent of preconditioner).
- Conferences promote work that wouldn't otherwise happen (thanks Gary). Shame when results are hopelessly incomplete.
- Wish `gdb` would run on my iMac. Use SuiteSparseQR mex for better access to UMFPACK, SPQR.
- LU preconditioning with U as right-preconditioner: can be slow. TPP might be OK for full-rank A, probably need TRP for singular A.